Chapter 7 Results

The quantitative data analysis will be broken into several parts; one for each of the primary factors measured. Many of these factors are derived from others. The actual source of the data or the procedure for computing the processed data will be described in each section. The rudimentary data includes search execution times (in decimal minutes) and position/orientation samples. The raw data is shown in Appendix C.

No one factor alone can describe the performance difference between treatments. However, taken collectively, the data presented here will show that a causal relationship exists between the wayfinding principles and task performance. Nevertheless, it cannot fully explain the nature of the causal relationship. This lies in the cognitive processes of the subjects during task execution. This will be discussed in Chapter 8.

For each factor, a Friedman two-way ANOVA was performed to determine the statistical significance of differences between treatments. This nonparametric test was chosen as the best method to measure differences in a small sample population using a repeated measures design.

Total Time

The observed total time for any treatment was taken to be the elapsed time from the initial presentation of the stimulus (the beginning of the first naive search subtask) to either the successful completion of relocating the home target after all five successful naive searches or voluntary failure as indicated by the subject. In cases where the subject did not successfully complete the task, the total time measurement was adjusted. Each incomplete naive search was replaced with the largest time of any *successful* naive search rounded up to the nearest whole minute. For example, in the control treatment, the largest time for any subject to complete any naive search was 11.4 minutes. Therefore, the time inserted into all incomplete naive searches in the control treatment is 12 minutes. This same technique was used for primed searches as well. The sum of all actual and penalized times is the adjusted total time. This is the factor reported in this section.

Total time should coarsely measure the efficiency of search during any trial. It is therefore expected that total time for the map treatments will be lower than the grid treatment with the control treatment having the highest times of all. The differences in total time are shown to be significant (Friedman test statistic = 15.96, $p \leq 0.001$) with the control treatment having the highest sum of ranks of 37.0. The order of presentation was found to be insignificant ($F(3,24) = 1.418$, $p = 0.262$). There was also no significant difference due to gender within any treatment (Control, $F(1,8) = 3.933$, $p = 0.083$; Map, $F(1,8) = 1.287$, $p = 0.289$; Grid, $F(1,8) = 1.020$, $p = 0.342$, Map/Grid, $F(1,8) = 1.007$, $p = 0.345$). The data are shown in Figure 7-1. The individual times for each treatment within each subject are shown vertically. The averages across subjects are shown in the last column and are extended across the chart with dashed lines.

The application of this factor is confounded by the fact that there are a number of interrelating criteria by which to evaluate the effectiveness or performance of any one trial. The total time factor is affected by others such as average velocity and percentage of the area searched. A subject could have conducted an organized, oriented search which was plagued with what will be referred to as “bad luck” or simply looking in the wrong places. Nevertheless, the very large separation (around 10 minutes) of the control treatment from all others, clearly shown in the graph, indicates that subjects were more often disoriented in this treatment causing repetitious search behavior not prevalent in other treatments. Furthermore, the relatively close grouping of the other three treatments indicates that although the presence of the maps seems to improve performance, other factors will better illustrate their differences.

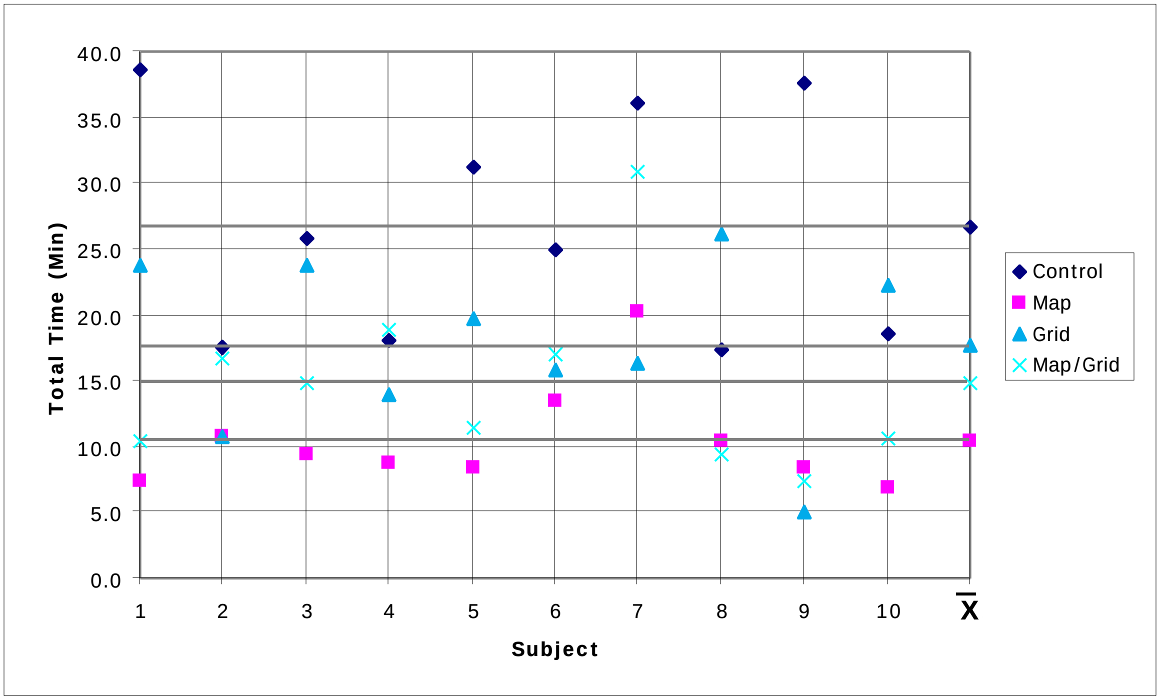


Figure 7-1 Total Time

Average Naive Search Time

In the case of average naive search time, the adjusted times as described in Total Time on page 115 are used along with actual times. The average naive search time for any treatment is simply the average time to completion of all successful and adjusted naive searches. It is necessary to include the adjusted times due to the fact that a large portion of the total time for many subjects was spent on unsuccessful searches. This time must be accounted for in some way.

Naive searches are dependent on the subject's ability to conduct an organized exhaustive search. Consequently, a measure of naive search time should show whether or not a subject was able conduct such a search or if disorientation caused the search to be ineffective. The differences in average naive search time were not found to be significant (Friedman test statistic = 6.84, $p = .077$) with the control treatment having the highest sum of ranks of 32.0. The order of presentation was found to be insignificant ($F(3,24) = 1.765$, $p = 0.181$). There was also no significant difference due to gender within any treatment (Control, $F(1,8) = 4.015$, $p = 0.080$; Map, $F(1,8) = 1.318$, $p = 0.284$; Grid, $F(1,8) = 3.112$, $p = 0.115$, Map/Grid, $F(1,8) = 1.086$, $p = 0.328$). See Figure 7-2.

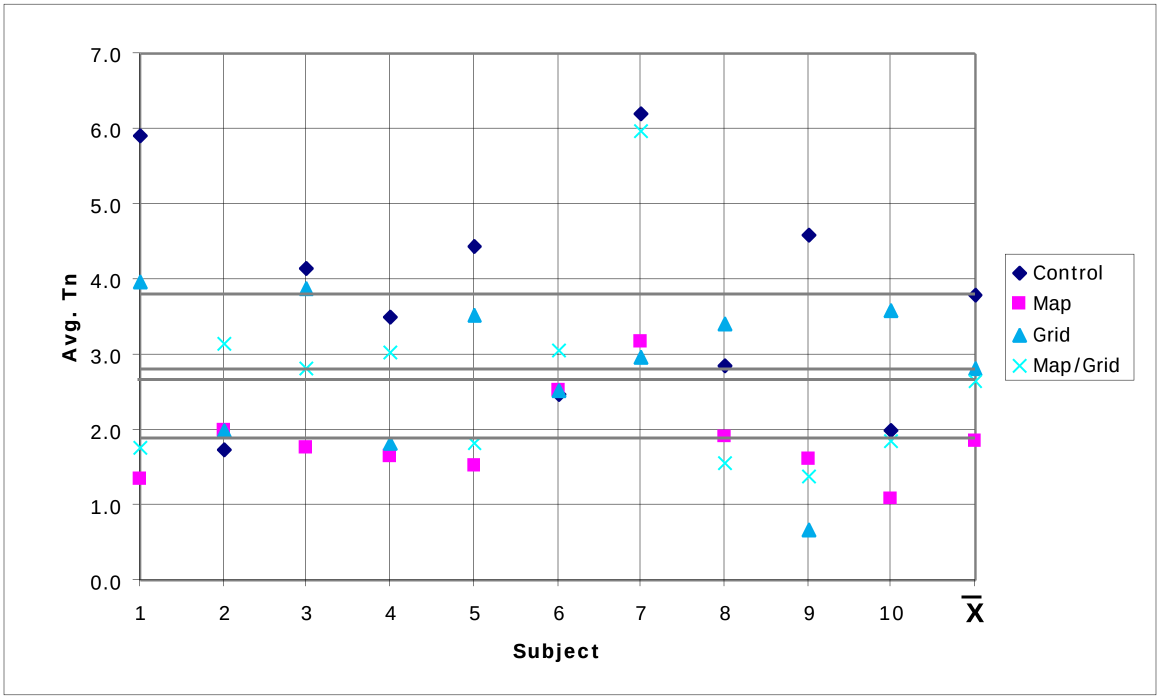


Figure 7-2 Average Naive Search Time

The predominance of naive searching over primed searching within the overall task causes the graph to mirror that of total time. However, the samples are more tightly grouped. We conclude that the primed search separates the total time samples more than the naive search. As with total time, a measure of time alone for a search task is incomplete in its ability to describe performance. Many of the search times, particularly in the grid treatment, were inflated due to a large amount of spatial information maintenance behavior (See Chapter 8).

Primed Search Time

The actual primed search time was taken to be the total time beginning immediately after the last successful naive search and ending with the completion of the primed search. If the primed search was unsuccessful, the time was adjusted as described in Total Time on page 115.

A proficient primed search requires an accurate knowledge of the environment and more importantly, of a precise path or method of locating the target. Consequently, a subject in any treatment conducting an efficient primed search will have a lower search time than a subject who either becomes disoriented or cannot properly direct the search. The differences in primed search time were found to be significant (Friedman test statistic = 13.56, $p \leq .005$) with the control treatment having the highest sum of ranks of 36.0. The order of presentation was found to be insignificant ($F(3,24) = 0.101$, $p = 0.958$). There was also no significant difference due to gender within any treatment (Control, $F(1,8) = 0.001$, $p = 0.976$; Map, $F(1,8) = 0.645$, $p = 0.445$; Grid, $F(1,8) = 0.630$, $p = 0.450$, Map/Grid, $F(1,8) = 0.073$, $p = 0.794$).

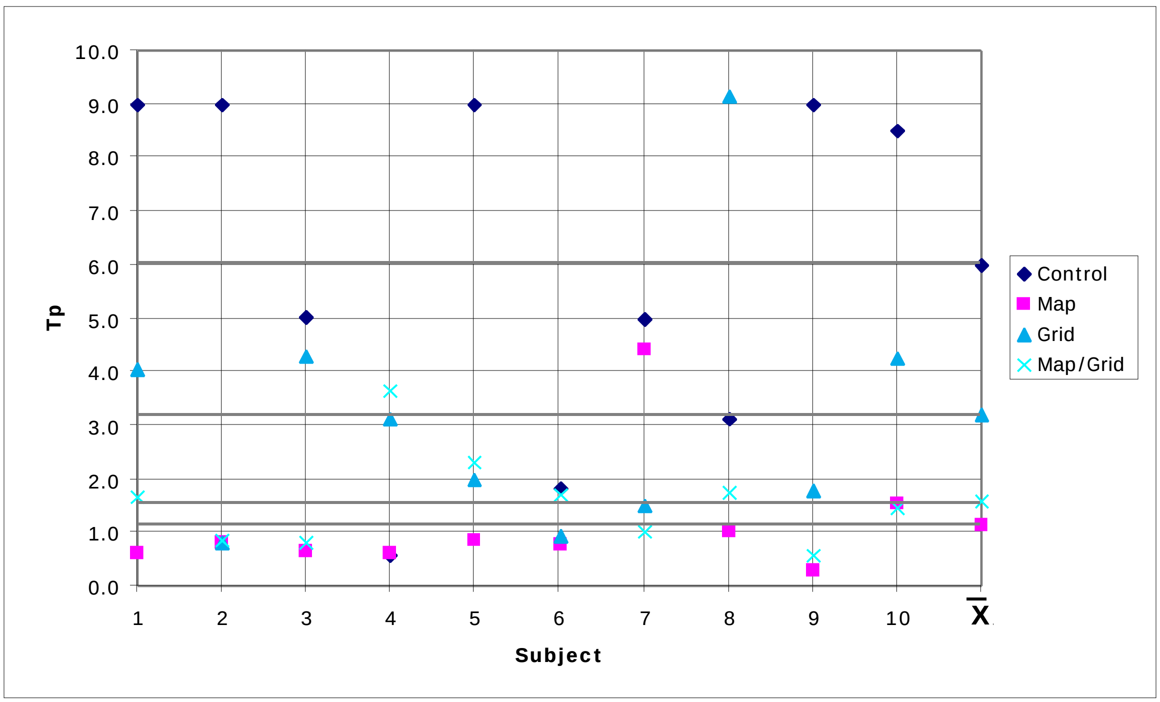


Figure 7-3 Primed Search Time

As with the total time factor, the control treatment is separated by a large margin (nearly 3 minutes). Also note that the two map treatments are clustered together with very low search times. The provision of global information within these treatments eliminates the need of the subject to infer directional information from the environment itself. Using the map (in either the map or the map/grid treatment), subjects were rarely disoriented and located the home target without wasted effort. The control treatment, however, often illustrated how ineffective a disoriented search can be. Most commonly, the subject was so disoriented at the inception of the primed search that it was effectively downgraded to another naive search requiring an exhaustive method. The grid treatment, while requiring more time than the map treatments, allowed subjects to maintain their orientation and consequently to infer direction to complete the search.

Total Distance Travelled

Total distance was determined by summing the distance between all sampled points taken during the trial. It is measured in kilometers. This method yields an accurate measure of distance travelled over the course of a complete trial. However, as is the case with total time, it must be adjusted to account for trials which were not successfully completed. This adjustment was done similarly to total time by adding a distance proportional to the adjusted time. A velocity is computed from the actual time and the actual distance. This is then multiplied by the adjusted total time to determine the adjusted total distance. This is the factor described in this section.

This factor complements the time factors in that it gives an indication of how much movement occurred. Accordingly, it should be considered along with the average velocity and percent of environment viewed factors in order to paint a more complete picture of movement and its structure. The differences in distance travelled were found to be significant (Friedman test statistic = 17.40, $p \leq .001$) with the control treatment having the highest sum of ranks of 37.0. The order of presentation was found to be insignificant ($F(3,24) = 0.971$, $p = 0.423$). There was also no significant difference due to gender within any treatment (Control, $F(1,8) = 0.024$, $p = 0.881$; Map, $F(1,8) = 0.075$, $p = 0.791$; Grid, $F(1,8) = 0.039$, $p = 0.848$, Map/Grid, $F(1,8) = 1.029$, $p = 0.340$). See Figure 7-4.

The large separation of the control treatment from the other three treatments (approximately 800km) indicates an exceptional amount of multiple path traversal. Subjects in the control treatment commonly searched the same space many times due to disorientation.

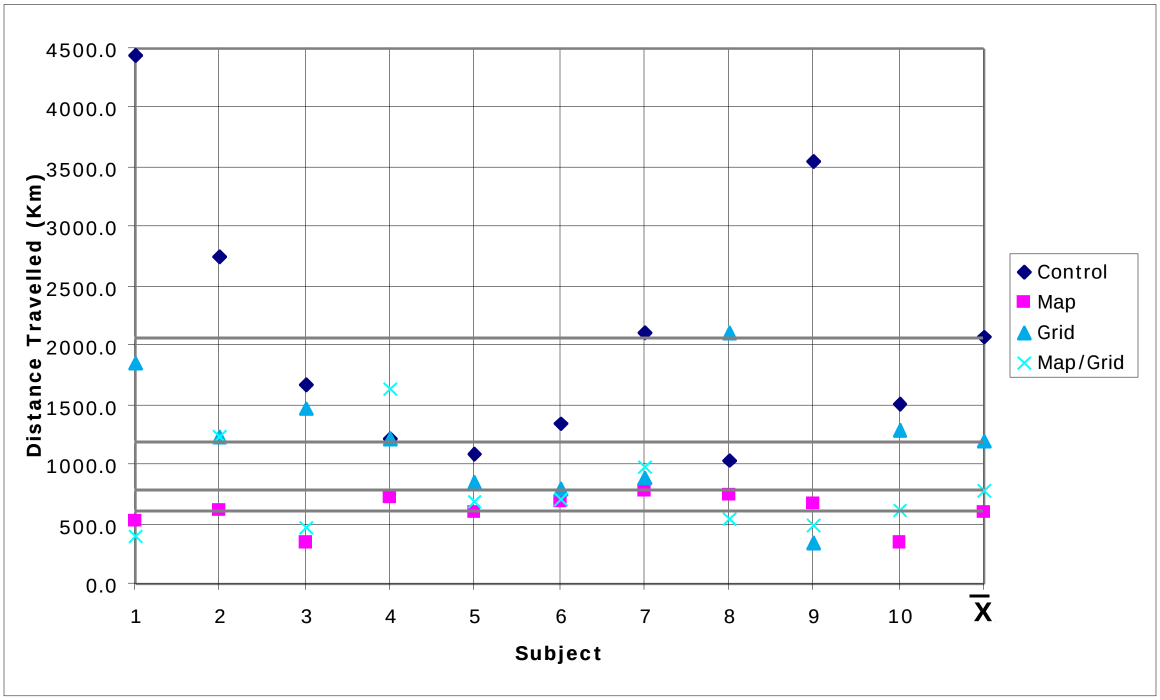


Figure 7-4 Distance Travelled

The larger values in the grid treatment over the map treatments is attributed to movements intended to maintain orientation. These movements were typically coordinated with the grid. The lower distances associated with the map treatments illustrates the ability of subjects within these treatments to accurately guide the search with little wasted movement. Behaviors such as heuristic searching were most commonly used within a map treatment.

Percent of Environment Viewed

Combining the sampled viewpoint positions and orientations with the field of view and view depth, a “footprint” of the volume of space viewed at every point was placed on the environment. Figure 7-5 shows the unsearched area, the unsearched land, and the searched area in shades of gray. The composite image was then processed to determine how much of the environment’s searchable space (the sea only in this case; the black area in the figure) remained unsearched. This was done with an image processing technique tabulating the total number of blue pixels and determining the overall percentage which is reported here.

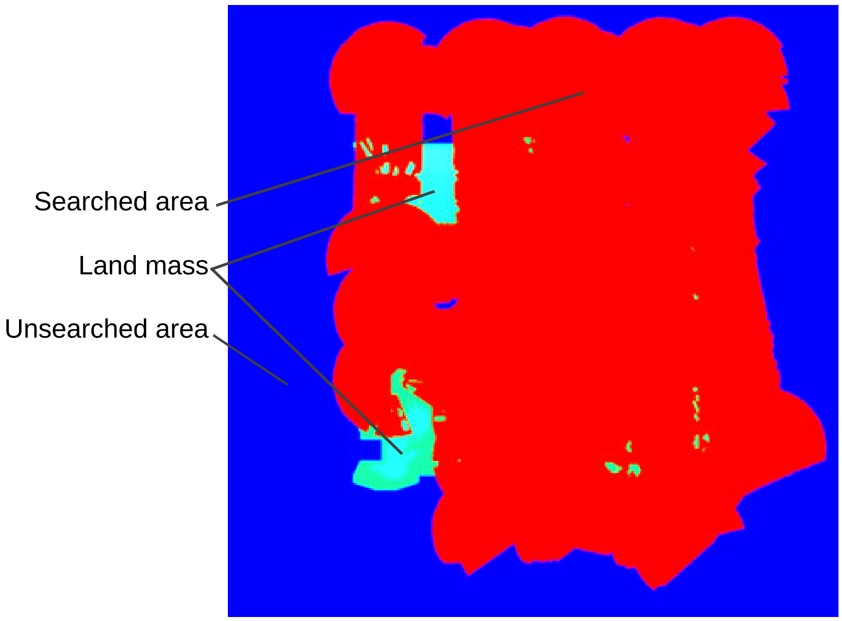


Figure 7-5 The percent of environment viewed method of measurement.

This factor differs from the distance travelled factor because it eliminates cases where the subject searched the same space multiple times. The differences in percent of environment viewed were found to be insignificant (Friedman test statistic = 3.00, $p = .392$) with the control treatment having the highest sum of ranks of 31.0. The order of presentation was also found to be insignificant ($F(3,24) = 1.935$, $p = 0.151$). There was no significant difference due to gender within any treatment (Control, $F(1,8) = 1.329$, $p = 0.282$; Map, $F(1,8) = 0.298$, $p = 0.600$; Grid, $F(1,8) = 0.003$, $p = 0.955$, Map/Grid, $F(1,8) = 0.166$, $p = 0.694$). See Figure 7-6.

The reason for this factor not showing significance is simply that "search efficiency" cannot be attributed to the amount of space searched without considering the time it took to search that space. Furthermore, what characterizes an efficient search? If the measure of performance depends on speed only, then time is the primary factor. However, if it depends on the volume of space searched, then the percent viewed factor is the primary factor. A more appropriate measure is the ratio of percent viewed to total time. For this measure, the differences between treatments were found to be significant (Friedman test

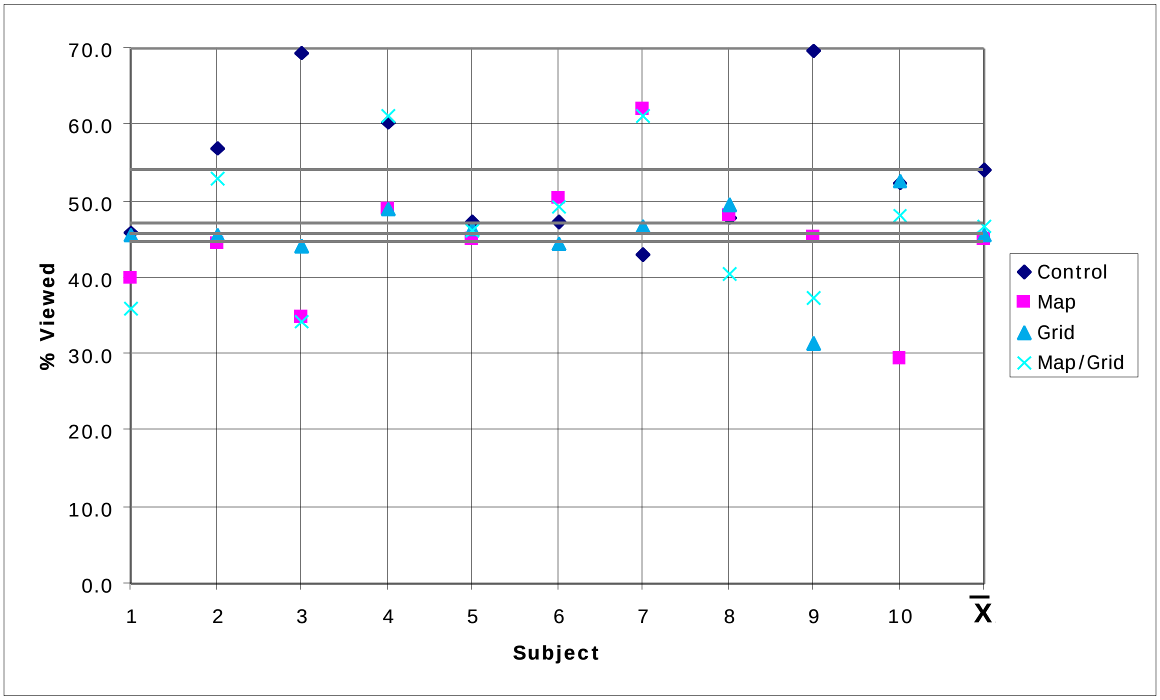


Figure 7-6 Percent of Environment Viewed

statistic = 13.80, $p \leq .005$) with the map treatment having the highest sum of ranks of 37.0. See Figure 7-6.

Those cases where large regions of space were searched in a relatively short amount of time (a high ratio on the graph) are more consistent with an intuitive definition of an efficient search. But again, this measure tends to penalize those subjects who conducted an effective search which, undoubtedly with some luck, located all targets without searching the entire environment and even more so, those subjects who happened to move relatively slowly.

Average Velocity

The average velocity is computed as the total distance divided by the total time. The differences in velocity were found to be significant (Friedman test statistic = 8.40, $p \leq .05$) with the control treatment having the highest sum of ranks of 32.0. The order of presentation was found to be insignificant ($F(3,24) = 0.935$, $p = 0.439$). There was also no significant difference due to gender within any treatment (Control, $F(1,8) = 1.460$, $p = 0.261$;

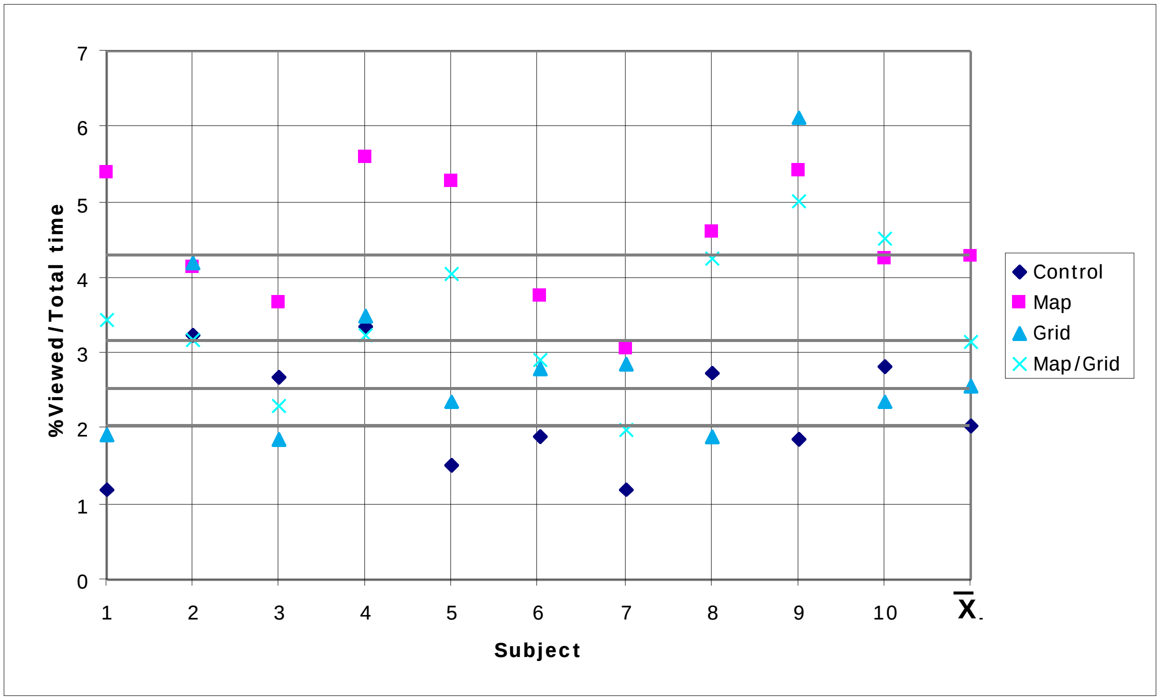


Figure 7-7 The Ratio of Percent Viewed to Total Time

Map, $F(1,8) = 1.469$, $p = 0.260$; Grid, $F(1,8) = 4.575$, $p = 0.065$) except the map/grid treatment ($F(1,8) = 12.917$, $p \leq 0.01$). See Figure 7-6.

The only other supporting evidence for this gender difference lies in the cognitive factors scores; specifically in the embedded figures test which also showed a difference. In general, females were more field dependent than males and they tended to move more slowly on the map/grid treatment. This could be explained by the similarities between the visual stimulus of the map with the grid and a complex figure as described in Appendix A. A field dependent subject would tend to have difficulty separating the grid from the environment and that difficulty could consequently cause the subject to move more slowly during searching tasks.

In general, velocity shows a slight tendency to be higher for the treatments which do not use a map. This can be accounted for by noting that subjects often will navigate by the map alone consulting the actual environment only when necessary. The analysis in the next chapter will show that more cognitive overhead went into the non-control treatments

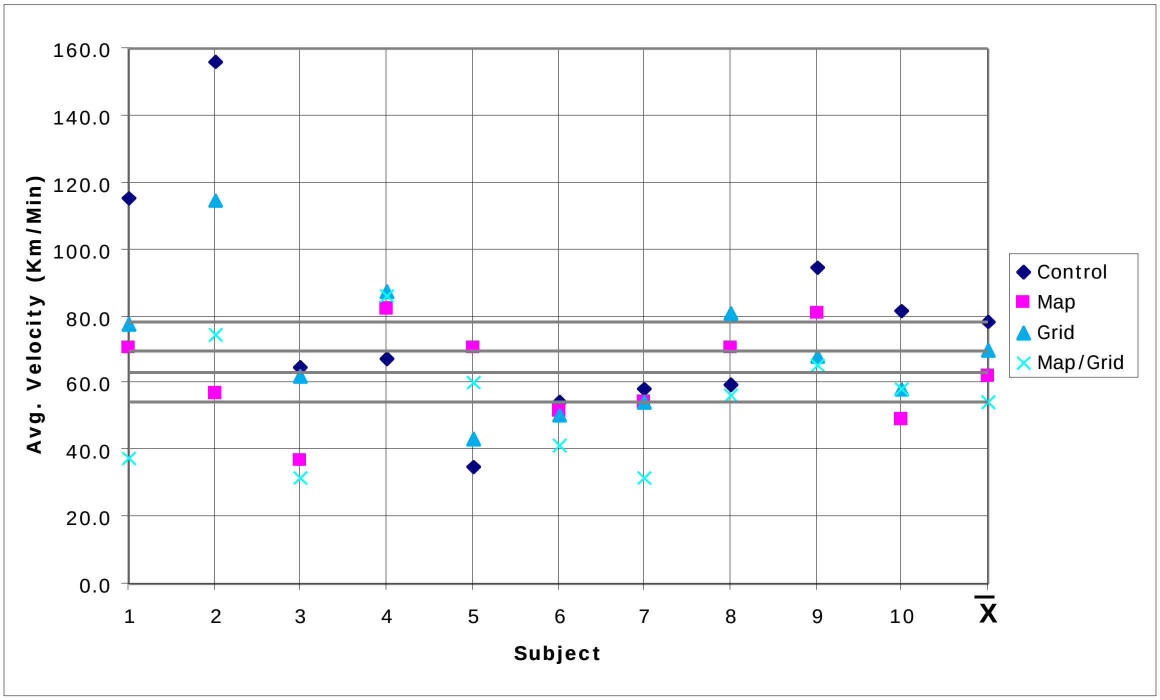


Figure 7-8 Average Velocity

due to the added stimuli. While it may seem that this would be a cause for slower performance in these treatments, it is important to note that this overhead enabled efficient search techniques thereby lowering time.

Map Distance Error

Map distance and direction error were determined using a metric intended to normalize results to make comparisons across subjects. The technique begins with measurements taken from an actual top-down view of each environment. A grid is placed over the world and the Cartesian coordinates of each target are marked. This is the left map in Figure 7-9. From this information, the distances and relative directions[†] between any two targets can be determined. These values are shown in the **Actual** columns in the table. The distance values (D) are given in centimeters and the direction angles are in degrees from zero. For consistency, one direction is chosen to be bearing zero for all cases and all maps (zero is

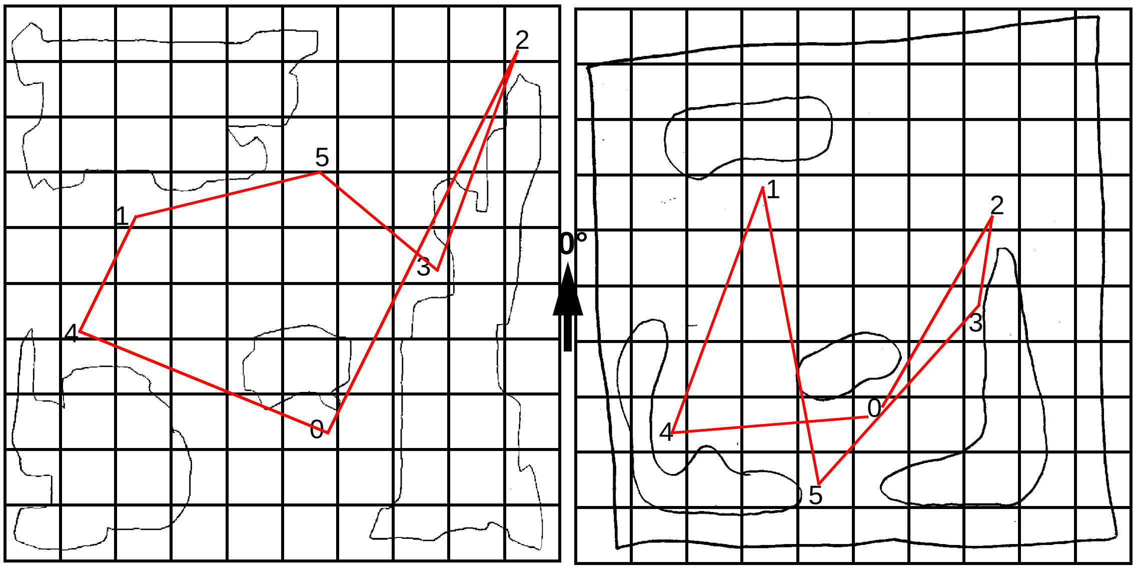
† To the resolution of the grid. A 1cm^2 grid was used in this analysis.

straight up in this example). Each map is evaluated similarly. The targets are identified by their Cartesian coordinates. The grids in Figure 7-9 have been added to show the method.

A path is constructed from the home target to the last target identified and back to the home target. In this case the ordering is 0-2-3-5-1-4-0 (Shown by a line in the figure). The distance and relative bearing of each leg of this path (i.e. target to target) is compared to the actual values. The distances must be normalized by using the overall distance of the entire path as one. For example, the leg 0-2 is measured to be 6.1cm on the subject’s map. The total path length is 34.4. Therefore the normalized 0-2 path length is 6.1 ÷ 34.4 or .177. These values are compared to those of the actual world resulting in the actual percentage error. In this case, the result is |**Norm Dist M - Norm Dist A**| or .122. This is shown in column **%ΔDist** and is averaged at the bottom of the column as the average distance error. The raw data for these calculations is included in Appendix F. Map drawings are shown in Appendix E.

The differences in map distance error were found to be insignificant (Friedman test statistic = 6.93, $p > .05$) with the control treatment having the highest sum of ranks of 33.0. The order of presentation was found to be insignificant ($F(3,24) = 1.835$, $p = 0.168$). There was no significant difference due to gender within any treatment (Control, $F(1,8) = 0.001$, $p = 0.977$; Map, $F(1,8) = 4.292$, $p = 0.072$; Grid, $F(1,8) = 2.840$, $p = 0.130$; Map/Grid, $F(1,8) = 0.041$, $p = 0.845$). See Figure 7-10.

Natural human abilities to judge distance and direction (Hale & Dittmar, 1994; Marshak, Kuperman, Ramsey & Wilson, 1987) show that distance estimation is typically easier than direction estimation. This is supported by the data here in that direction errors are of an order of magnitude greater than distance errors. Furthermore, the fact that distance error did not prove to be significantly different across treatments is not surprising in light of this ability. The control treatment is clearly separated from the other three treatments which are clustered tightly together. However, the separation is only a 2% difference. The variance is very high as there is no clear treatment in which distance estimation was significantly easier than another. We can only conclude from this that the cues for determining



S4	Control		Norm Dist M	Actual		Norm Dist A	%Δ Dist	%Δ Angle
0	**D**	θ		**D**	θ			
2	6.1	35	0.177	12.5	28.6	0.3	0.122	0.036
3	2.7	201.8	0.078	6.5	202.6	0.156	0.077	0.004
5	6	228.4	0.174	4.9	315	0.118	0.057	0.481
1	7.2	347.9	0.209	5.8	250	0.139	0.07	0.544
4	6.8	197.1	0.198	4.7	198.4	0.113	0.085	0.007
0	5.6	79.7	0.163	7.3	105.9	0.175	0.012	0.146
Σd	34.4			41.7			0.071	0.203

Figure 7-9 An example of the map distance and direction metric technique.

motion, specifically the sea terrain and the frame rate (optical flow), were effective in providing the necessary information to infer distance. However, the same cannot be said of direction.

Map Direction Error

Similarly to the map distance error metric, map direction error is also measured in percent error from the original world. In the example of Figure 7-9, the relative bearing from target 0 to target 2 is measured to be 35° on the drawn map and 28.6° in the actual world. These can be compared directly as no normalization is necessary. The maximum error is taken to be 180° which would place a target on the exact opposite side of its correct orien-

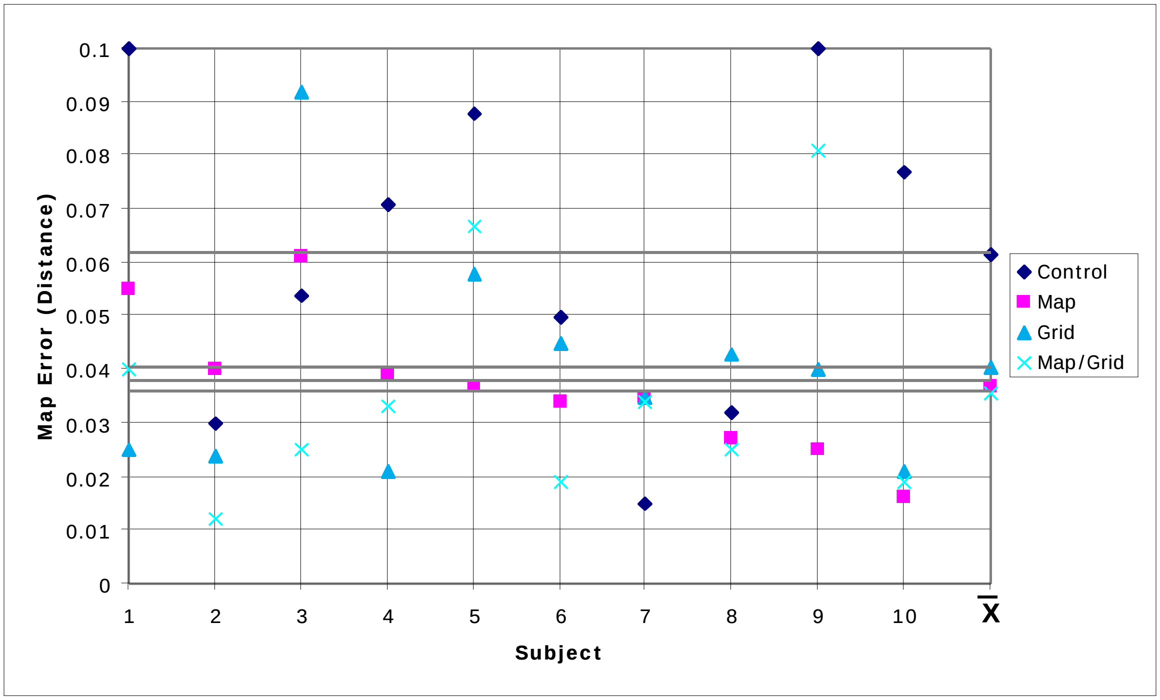


Figure 7-10 Map Distance Error

tation. Therefore, the percent error is computed as the difference in measured angle divided by 180°. In this case, (35 - 28.6) ÷ 180 = .036. As with the distance errors, these are tabulated and averaged at the bottom of column **%ΔAngle** as the average direction error for this map.

The differences in map direction error were found to be significant (Friedman test statistic = 14.52, $p \leq .005$) with the control treatment having the highest sum of ranks of 37.0. The order of presentation was found to be insignificant ($F(3,24) = 0.246$, $p = 0.863$). There was no significant difference due to gender within any treatment (Control, $F(1,8) = 0.592$, $p = 0.464$; Map, $F(1,8) = 0.786$, $p =0.401$; Grid, $F(1,8) = 0.762$, $p = 0.408$; Map/Grid, $F(1,8) = 0.012$, $p = 0.916$). See Figure 7-11.

Again, the control treatment is separated from the other treatments by a large margin (nearly a 15% difference). This is expected considering how much of the search time during control treatments found subjects disoriented. What is not necessarily expected is the low error of the grid and map/grid treatments but not the map treatment. Clearly, the presence of the grid as an absolute orientation reference affected the ability of subjects to place

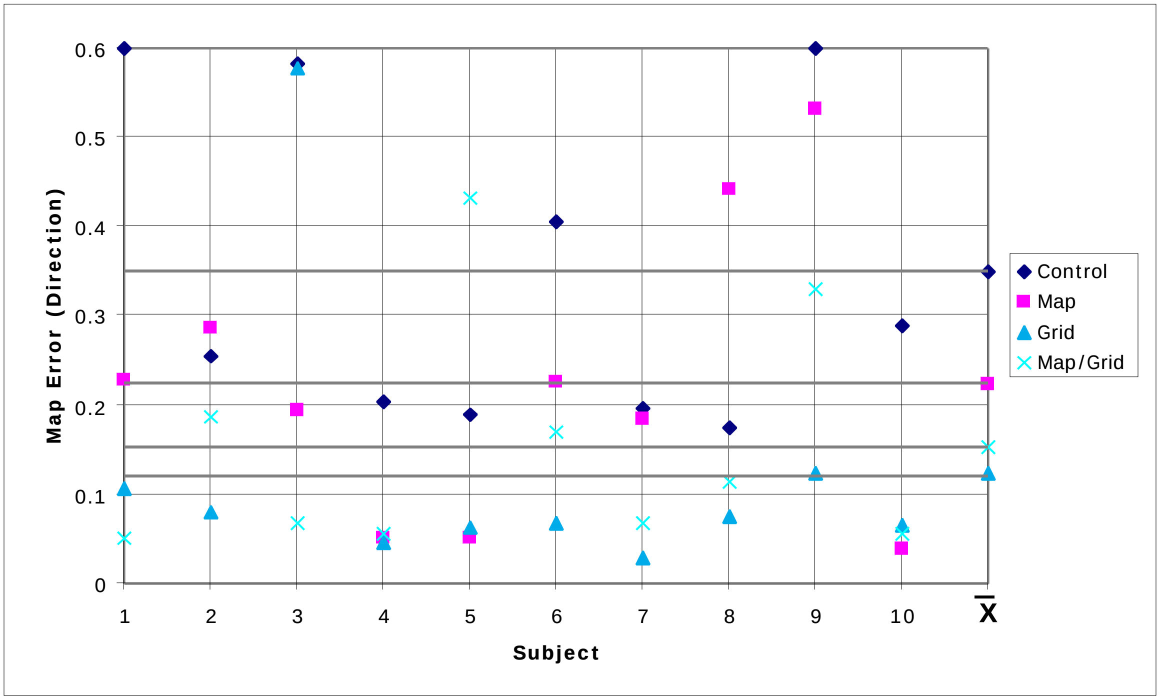


Figure 7-11 Map Direction Error

targets on the map. Furthermore, the extra attention given the grid during the grid treatment is evident in its overall superior error percentage. This is not true when the land masses are considered.

Land Map Error

There was significant difficulty in defining a practical, yet descriptive metric for the land masses on the maps. The first impulse was to use a "best fit" procedural method which would match random points on the drawn map land masses to their corresponding points on the actual land masses. This was intended to be closely similar to the metric used for target distance and direction. However, there are two fundamental problems with this approach; 1) how does one choose the points to use and subsequently, 2) how does one determine where its counterpart is in the actual world? Unlike the target distance and direction metric, this is a pattern recognition problem. We want to know which maps show the land masses most similarly to the actual world. This includes the land masses' shape, relative size, position, and orientation. It was determined that the best way to perform such a complex pattern recognition task was to allow a number of people to subjectively rate the maps in terms of the attributes previously mentioned. Furthermore, rather than use the opinion of the average person who may or may not have a proficient internal metric by which to judge a map, the evaluations of four geographers from Eastern Michigan University were used. The instructions given them are reproduced in Appendix D.

The resulting data is a rank ordering of the maps for each world which were then separated by treatment. The differences in land representation were found to be highly significant (Friedman test statistic = 18.84, $p \leq .0001$) with the grid treatment having the highest sum of ranks of 34.0. The order of presentation was found to be insignificant ($F(3,24) = 0.345$, $p = 0.793$). There was no significant difference due to gender within any treatment (Control, $F(1,8) = 5.448$, $p = 0.048$; Map, $F(1,8) = 0.478$, $p =0.509$; Grid, $F(1,8) = 1.505$, $p = 0.255$; Map/Grid, $F(1,8) = 0.127$, $p = 0.730$). See Figure 7-12.

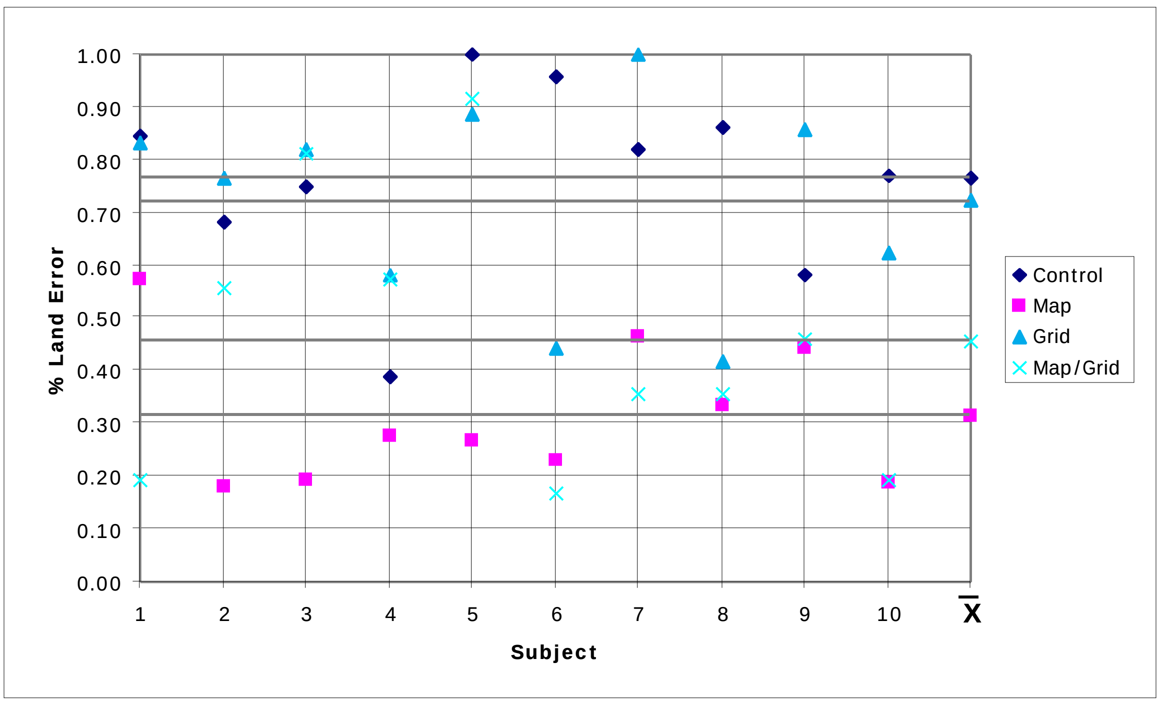


Figure 7-12 Map Land Error

As expected the two map treatments which provide a geocentric view of the environment are clearly separated from the two egocentric treatments (control and grid). Targets are thought of as points which can be placed relative to the grid resulting in accurate map drawings. The shape and size of land masses must be inferred in the egocentric treatments and consequently is more poorly represented in the grid treatment than in the map treatments.

Total Map Error

Total map error is computed as the sum of all three map error computations; map distance error, map direction error, and land error. All three have been computed in terms of a percentage. The differences in map error were found to be significant (Friedman test statistic = 17.40, $p \leq .001$) with the control treatment having the highest sum of ranks of 37.0. This is expected since a simple sum of errors without artificial weighting will concur with the greatest contributors; mainly direction and land error. Distance error, which was the map error factor with least significance, is also the least contributor having error rates under 10%.

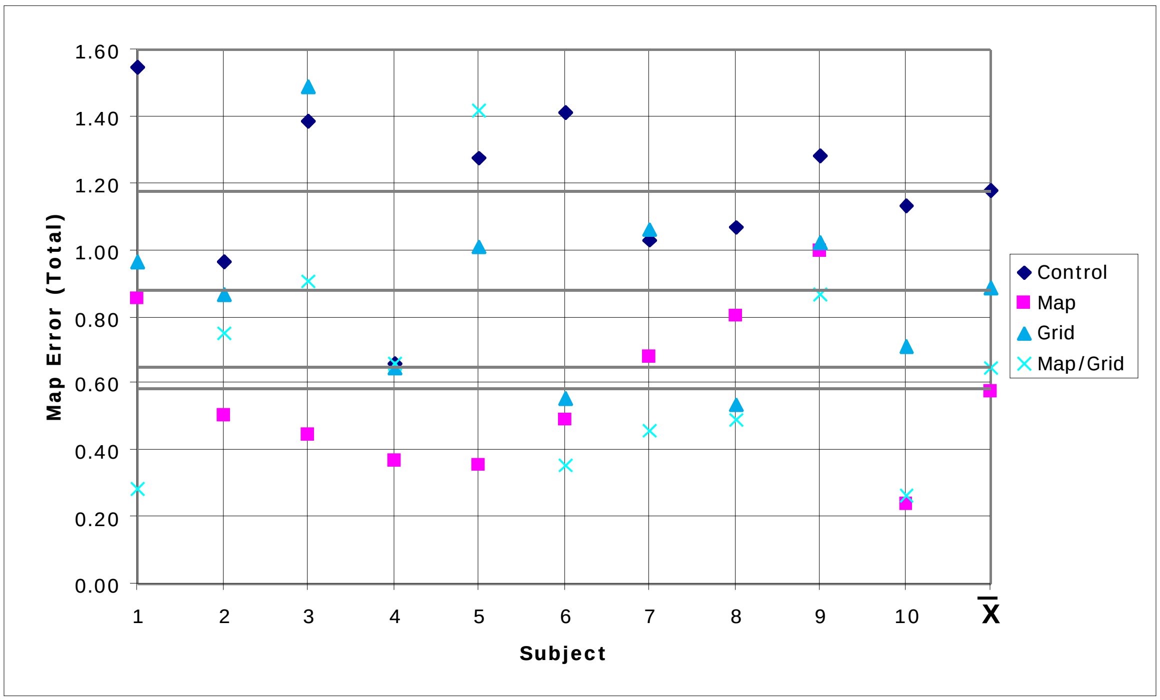


Figure 7-13 Total Map Error

Cognitive Factors Examinations

The cognitive factors examinations are standardized tests. The results reported in this study can be compared to the established population averages as reported in Ekstrom, et al. (1976) and Witkin, et al. (1971). However, the majority of the published data is from the 1963 version of the tests making direct numerical comparisons difficult.

The purpose in studying cognitive factors within the context of this experiment was to investigate relationships between natural human spatial ability and wayfinding performance. The results showed a generally weak correlation to wayfinding (For details see Chapter 9). The actual numerical data is presented in Appendix C. Contrary to expected results (Ekstrom, et al., 1976; McGee, 1979) but similar to the wayfinding factors discussed earlier in this chapter, cognitive factors scores in spatial orientation ($F(1,8) = 1.432$, $p > 0.266$) and visualization ($F(1,8) = 0.229$, $p > 0.645$) did not show significant gender differences. However spatial scanning ($F(1,8) = 6.621$, $p \leq .05$) and embedded figures ($F(1,8) = 6.710$, $p \leq .05$) did show a significant gender difference. See Figure 7-14.

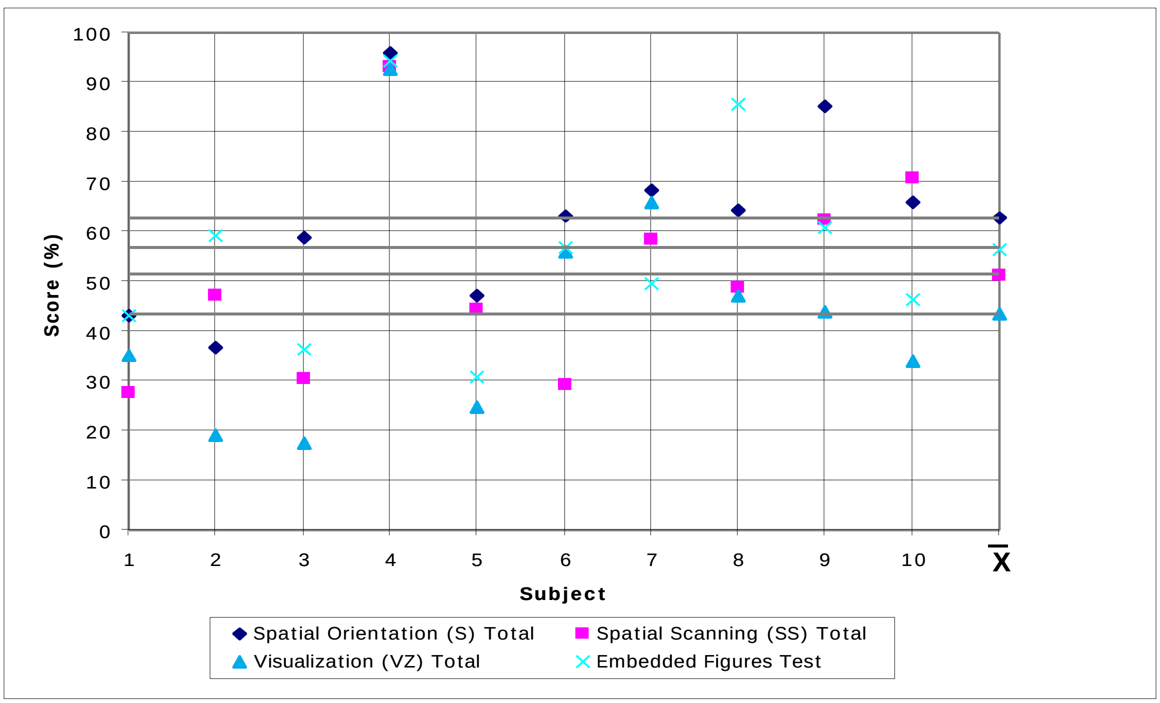


Figure 7-14 Cognitive Factors

The lack of gender difference in spatial orientation and visualization can be attributed to the subject pool used in this experiment. All subjects were of a technical background. The greatest gender difference was in the Embedded Figures Test in which males tended to be less field dependent than females. This difference has been discussed in Average Velocity on page 123. Cognitive factors scores will be discussed further in Chapter 9.

Power Analysis

A post-hoc power analysis of the Friedman two-way analysis of variance by ranks found power to be 0.2330 with N = 10 and $\alpha = 0.05$ for a large effect size. Similarly, a post-hoc power analysis of the four-way ANOVA used to measure treatment order significance found power to be 0.2353 with N = 10 and $\alpha = 0.05$ for a large effect size. Lastly, a post-hoc power analysis of the Pearson correlations used to measure correlations between factors (See Chapter 9 for results) found power to be 0.3628 with N = 10 and $\alpha = 0.05$ for a large effect size.

The high power values for all tests indicates that for each hypothesis under evaluation, there is a relatively high probability of producing a Type II error. That is, the null hypothesis H_o may not have been rejected when it was, in fact, false. A larger number of subjects would be needed to lower the power values to the necessary level.

Thus far, the quantitative results have presented strong evidence supporting the hypothesis that the presence of the wayfinding augmentations did significantly increase searching performance. Furthermore, in general, it has been shown that the two map treatments have a stronger effect than the grid treatment but that all three are significantly better than the control treatment. However, little has been presented to show any difference between the two map treatments nor has any qualitative data been presented to describe the actual observed differences. This is the focus of the next chapter.